Isis[1]: Protecting Children in Online Social Networks

Awais Rashid, Paul Rayson, Phil Greenwood, James Walkerdine
Computing Department, Lancaster University, UK
{awais | greenwop | paul | walkerdi} @comp.lancs.ac.uk
Penny Duquenoy, Patrick Watson
Middlesex University, UK
P.Duquenoy@mdx.ac.uk, P.Watson@mdx.ac.uk
Margaret Brennan
Child Exploitation and Online Protection Centre, UK
maggie.brennan@ceop.gov.uk
Matt Jones
Swansea University, UK
mattjonez@gmail.com

Abstract

The aim of the Isis project is to develop an ethics-centred monitoring framework and tools for supporting law enforcement agencies in policing online social networks for the purpose of protecting children. The project is developing natural language analysis techniques to help identify child sex offenders from chat logs and monitoring mechanisms that can be non-invasively attached to file sharing systems for identifying the distributors of child abuse media. The ethical issues associated with such monitoring activities are studied through consultations with representatives from stakeholder communities and fed back into the development of the framework and tools. The project results are to be used and evaluated by specialist international law enforcement agencies as part of their own policing activities.

1. Introduction and Overview

Recent years have seen a rapid rise in the number and use of online social networks. Such social networks vary in nature from chat systems, for example, MSN, Skype and IRC, to online communities, such as, MySpace and YouTube, through to file sharing systems, for instance, peer-to-peer networks: Gnutella, BitTorrent, FastTrack, etc. Amongst the many types of 'risk' on the internet as mentioned in the Byron review in the UK [1] and Internet Safety Technical Task Force in the US [2], these social networks pose two significant risks in terms of child exploitation. The first major type of risk is *paedophiles and other child sex offenders predating on children*. Children actively participate in social interactions using forums such as chat rooms and web-based communities. Offenders can use such forums to predate on children, or even to plan the commission of sexual offences against children. These concerns are reflected by the formation of the Virtual Global Taskforce and specialist UK enforcement agencies and Scottish legislation to criminalise the 'grooming' of children in chat rooms in October 2004. The second risk is the offence of *distributing and sharing child abuse media*. Child sex offenders can formulate their own social networks using mechanisms, such as file-sharing systems, in order to distribute and share child abuse media. The scale of distribution of illegal media (including child abuse media) on such file-sharing systems was highlighted by a recent study at Lancaster University [3], which found that 1.6% of searches and 2.4% of responses on the Gnutella peer-to-peer network relate to illegal sexual content. Given the system's scale, these results suggest that, on the Gnutella network alone, hundreds of searches for illegal images occur each second. The study also found that, of those users sharing illegal sexual content, 57% were solely devoted to such distribution while half of the material shared by another 17% involved such content.

[1] This research is supported by a UK research grant from EPSRC/ESRC (reference EP/F035438/1) involving the Universities of Lancaster, Middlesex and Swansea. For further details, see the project website (http://www.comp.lancs.ac.uk/isis/)

Given the vast amount of information that is communicated within online social networks, new monitoring and analysis technologies need to be developed in order to tackle the growing problem of child grooming and the distribution of child abuse media. The development of such technologies faces three significant research challenges:

1. *How to identify active child sex offenders across online communities?*
   Paedophiles and other child sex offenders often masquerade as children in order to establish contact with potential victims and gain their trust. Distinguishing the "innocent" interaction amongst children or amongst children and adults from such predatory advances is a non-trivial task yet effective, early and accurate identification of sexual offenders is vital for the protection of children. At the same time such offenders may use multiple online identities and known child sex offenders may move to other online social networks upon detection in one network. It is, therefore, vital that once a suspected child sex offender is detected in one network, s/he can be successfully detected in other networks which s/he may attempt to employ for grooming children.
2. *How to identify the core distributors of child abuse media?*
   The key research challenge is to accurately identify child abuse media from the plethora of perfectly legal material that exists within file sharing systems. The problem is compounded by the fact that offenders often use specialised vocabulary to describe their shared media—a vocabulary that evolves and changes over time—and operate over different file sharing networks. Any monitoring framework must be non-invasively attachable to existing file sharing systems given the wealth of such systems and clients available today. In addition to identifying child abuse media within such systems, any monitoring tools must be able to distinguish core distributors of such media from mere users. This is essential for child protection as this would help law enforcement agencies in tackling the problem at its roots.
3. *How to ensure that such developments maintain ethical practices?*
   The development of such monitoring and analysis techniques raises a number of ethical challenges pertaining, on the one hand, to utilising the framework and tools in a beneficial way for child protection and, on the other hand, the need to protect innocent users of online social networks from the potential of falsely being identified as child sex offenders and safeguarding their privacy.

Isis is aiming to tackle the above three challenges by developing novel chat log analysis and non-invasive file sharing monitoring techniques based on natural language processing and aspect-oriented programming [4] practices respectively. The resulting framework and tools will assist law enforcement agencies while ensuring that they fall within current ethical bounds—note our goal is not automation but to provide support for detecting potential sexual offences through analysis of large amounts of data which cannot manually be analysed in an efficient manner.

2. Challenges Tackled by the Isis Approach

Existing work on policing online social networks has focused primarily on the monitoring of chat and file sharing systems. Chat policing software for home use such as Spector Pro[2], Crisp[3] and SpyAgent[4] allow the logging of online conversations, but are restricted in that they need to be installed on the actual PC that is partaking in the activity. Less obtrusive chat policing systems, as used by policing organisations, typically use a network-level tracing methodology [5] to identify and log chat traffic at the network-level for later analysis.

[2] www.spectorsoft.com
[3] www.protectingeachother.com
[4] www.spytech-web.com

However, police surveillance tactics deployed at network-level present real challenges to law enforcement in terms of detecting edge-based criminal activity and achieving effective online guardianship [6]. Three significant shortcomings can be observed. Firstly, too much data is produced to make pro-active analysis practical. Secondly, child sex offenders often masquerade as children in order to make contact, making detection difficult. Thirdly, systems tend to be developed to monitor a predator stereotype (adult male) which does not reflect patterns of internet based sexual predation of children and young people [7]. For example, Finkelhor [8] found that young people themselves make aggressive sexual solicitations in almost half of all cases and that of those known to be adults (25%), the majority are aged between 18-25. In 27% of cases in this study (conducted in the US) the age of predators was unknown and could well include adults masquerading as young people. A key question yet to be addressed is how to distinguish both between adults predating as young people and between 'normal' youth sexual behaviour on the internet and youth predation. Due to these challenges, policing organisations focus primarily upon reactive policing, wherein known culprits are identified and tracked and children are provided with mechanisms to report suspicious behaviour. Unfortunately, this approach is incapable of tackling many cases where children do not report incidents (in [8] only 3% reported) and where offenders may be unknown to the authorities. Moreover, these policing tactics do little to advance a preventive approach to the problem of online grooming and predation within social networks by enabling effective guardianship and the potential for law enforcement intervention in pre-criminal situations, e.g., at the point of an early "friendly" online encounter between a prospective offender and a child (see [6] for a discussion of the significance of offender search, pre-criminal situations, opportunity and other contextual factors in the prevention of Internet crimes against children).

In terms of language monitoring capabilities the existing chat policing software tools rely on human monitoring of logs or simple-minded keyword or phrase detection based on user-defined lists. Such techniques do not scale. Nor do they enable identification of adults masquerading as children or support pro-active policing. Techniques do exist which make use of statistical methods from computational linguistics and corpus-based natural language processing to explore differences in language vocabulary and style related to age of the speaker or writer. The existing methodologies, such as key word profiling [9], draw on large bodies of naturally occurring language data known as *corpora* (sing. *corpus*). These techniques already have high accuracy and are robust across a number of domains (topics) and registers (spoken and written language) but have not been applied until now to uncover deliberate deception. The second relevant set of techniques is that of authorship attribution. The current methods [10] would allow a narrowing in focus from the text to the individual writer in order to generate a stylistic fingerprint for authors.

For policing file sharing systems two significant tools exist, Peer Precision[5] and LogP2P[6]. Both systems also use a network-level tracing methodology in conjunction with a 'honey-pot' approach, wherein the policing peer offers an illegal file to the network and when an offender attempts to download this file, client-side software captures the offender's IP address at the packet-level. This approach suffers from two significant shortcomings. Firstly, it is unable to differentiate between those who download and share a single file, and those who are the 'core' distributors of child abuse media (e.g. distributing many thousands of files, producing and distributing child abuse imagery or uploading newly-produced child abuse material for the first time). This is a significant problem for frequently backlogged child protection agencies with limited resources. Secondly, as these systems work at the network-level, they can potentially be thwarted by encryption at the application-level. This is of particular significance as recent research has shown that users are migrating to more anonymous and secure file sharing systems [3]. Finally, and perhaps most critically, the honey-pot approach

[5] www.icactraining.org/P2P.htm
[6] aidounix.com/?LogP2P

relies upon the use of well-known files. Hence, it is incapable of identifying those offenders who may be sharing recently-produced material. The incorporation of monitoring functionality in file sharing systems requires significantly altering multiple components to ensure that monitoring takes place at the right points in the system. However, such invasive changes are expensive and hard to maintain and evolve across various releases of a system. The recent rise of aspect-oriented software development techniques [4] has facilitated non-invasive composition of such systemic concerns as monitoring, which makes in-step evolution of such functionality with changes in the rest of the system more modular and manageable. Though aspect-oriented techniques have been used in individual systems (e.g., the widely used mysql database system) for logging purposes, to date, they have neither been applied for monitoring online social networks nor on a scale spanning multiple systems and various releases of such systems. A particular issue in file sharing systems is that filenames reflect specialised vocabulary which changes over time [11].

Taking an ethical perspective on this research is not only important in respect of the abuse of children [12], and the protection of the researchers who conduct this type of research, but also because of the issues surrounding the use of monitoring technologies that have an impact on user privacy [13]. Researchers in the field of computer ethics have noted that values are embedded within technology design, e.g., [14, 15], and, as a result, there have been numerous calls for the integration of ethical assessment, evaluation and stakeholder impact analysis within the design and development of computer systems to mitigate adverse effects, e.g., [16, 17]. In advocating this approach there is a recognition that not only are the potential risks associated with the software development reduced [18], but also that the awareness of the development team to the ethical aspects inherent in these systems is raised – thus creating a body of 'ethically aware' information professionals [6]. To date there has been a lack of suitable case studies in the computer ethics literature and appropriate guidance for technology developers to incorporate ethical considerations within the development cycle. This project is developing new understandings of user-centred methods for highly sensitive systems and of effective designs of privacy/awareness interfaces that will benefit other developments and mitigate the effects of adverse outcomes that impact on public acceptability.

References

[1] T. Byron, "Safer children in a digital world: the report of the Byron review", http://www.dcsf.gov.uk/byronreview/, 2008.

[2] J. Palfrey, "Enhancing child safety & online technologies: final report of the Internet safety technical task force", Berkman Center, Havard University, 2008.

[3] D. Hughes, S. Gibson, J. Walkerdine, G. Coulson, "Is deviant behaviour the norm on P2P file sharing networks?" *IEEE Distributed Systems Online* 7(2), 2006.

[4] R. Filman, T. Elrad, S. Clarke, M. Aksit (eds.), "Aspect-Oriented Software Development", Addison-Wesley, 2001

[5] D. Hughes, J. Walkerdine, K. Lee, "Monitoring challenges and approaches for P2P file sharing systems", Proc. 1st International Conference on Internet Surveillance and Protection (ICISP'06), 2006.

[6] M. Taylor, E. Quayle. "The Internet and Abuse Images of Children: Search, Pre-criminal Situations and Opportunity" *Situational Prevention of Child Sexual Abuse*, R. Wortley, S. Smallbone (eds.), Criminal Justice Press, 2006. pp. 169-195.

[7] S. Dombrowski, K. Gischlar, T. Durst, "Safeguarding young people from cyber pornography and cyber sexual predation: a major dilemma of the Internet". *Child Abuse Review* 16, 2007, pp. 153-170.

[8] D. Finkelhor, K. Mitchell, J. Wolak, "Online Victimization: A Report on the Nation's Youth", National Center for Missing and Exploited Children, Alexandria, VA, 2000.

[9] P. Rayson (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*. 13:4 pp. 519-549.

[10] P. Juola, J. Sofko, P. Brennan, "A prototype for authorship attribution studies", *Literary and Linguistic Computing* 21, 2006, pp. 169-178.
[11] D. Hughes, P. Rayson, J. Walkerdine, K. Lee, P. Greenwood, A. Rashid, C. May-Chahal, C., M. Brennan (2008) Supporting law enforcement in digital communities through natural language analysis. In *proceedings of the 2nd International Workshop on Computational Forensics (IWCF 2008)*, Washington DC, USA, August 7-8, 2008. Lecture Notes in Computer Science 5158, pp. 122-134.
[12] M. Eneman, "The new face of child pornography", in *Human Rights in the Digital Age*, Cavendish Publishing, 2005.
[13] D. J. Cook, S. K. Das, "How smart are our environments? An updated look at the state of the art", *Pervasive and Mobile Computing* 3(2), 2007, pp.53-73.
[14] H. Nissenbaum, "Values in the design of computer systems", *Computers and Society*, March, 1998.
[15] J. van den Hoven, "ICT and value sensitive design", in *The Information Society: Innovation, Legitimacy, Ethics and Democracy*, P. Duquenoy, P. Goujon, K. Kimppa, S. Lavelle (eds.), Springer, 2007.
[16] P. Duquenoy, O. Burmeister. "Exploring ethical aspects of Pervasive Computing" in *Risk Assessment and Management in Pervasive Computing: Operational, Legal, Ethical and Financial Perspectives*, Varuna Godara (Ed.), IGI Global. 2008. pp.264-284.
[17] D. H. Gleason, "A software development solution", *Proc. Ethicomp: Systems of the Information Society*, Poland, 2001.
[18] D. Gotterbarn, "Reducing software failures: Addressing the ethical risks of the software development lifecycle", Australian Journal of Information Systems, 1999.
[19] P. Duquenoy, D. Whitehouse, "A 21st century ethical debate: Pursuing perspectives on Ambient Intelligence", *Proc. Landscapes of ICT and Social Accountability*, Finland, 2005.